

Running head: DUNNING-KRUGER IN REASONING

Dunning-Kruger effects in reasoning: Theoretical implications of the failure to recognize incompetence

Gordon Pennycook^{1*}, Robert M. Ross^{2,3}, Derek J. Koehler², & Jonathan A. Fugelsang⁴

¹Department of Psychology, Yale University, New Haven, United States

²Department of Psychology, Royal Holloway, University of London, London, United Kingdom

³ARC Centre of Excellence in Cognition and its Disorders, Macquarie University, Sydney,
Australia

⁴Department of Psychology, University of Waterloo, Waterloo, Canada

*gordon.pennycook@yale.edu

Word count: Main text = 7,591

References: 54

Cite as: Pennycook, G., Ross, R.M., Koehler, D.J., & Fugelsang, J. (2017). Dunning-Kruger effects in reasoning: Theoretical implications of the failure to recognize incompetence. *Psychonomic Bulletin & Review*, 24, 1774-1784.

Abstract

The Dunning-Kruger effect refers to the observation that the incompetent are often ill-suited to recognize their incompetence. Here we investigate potential Dunning-Kruger effects in high-level reasoning and, in particular, focus on the relative effectiveness of metacognitive monitoring among particularly biased reasoners. Participants who made the greatest number of errors on the Cognitive Reflection Test (CRT) overestimated their performance on this test by a factor of more than three. Overestimation decreased as CRT performance increased and those who scored particularly high underestimated their performance. Evidence for this type of systematic miscalibration was also found on a self-report measure of analytic thinking disposition. Namely, genuinely non-analytic participants (based on CRT performance) over-reported their 'Need for Cognition' (NC), indicating that they were dispositionally analytic when their objective performance indicated otherwise. Furthermore, estimated CRT performance was just as strong a predictor of NC as was actual CRT performance. Our results provide evidence for Dunning-Kruger effects both in estimated performance on the CRT and in self-reported analytic thinking disposition. These findings indicate that part of the reason why people are biased is because they are either unaware of or indifferent to their own bias.

Dunning-Kruger effects in reasoning: Theoretical implications of the failure to recognize incompetence

If competency is required to recognize incompetence, truly incompetent people will be both incompetent and unaware of their incompetence (Dunning, Johnson, Ehrlinger, & Kruger, 2003; Kruger & Dunning, 1999). The failure to recognize incompetence among the incompetent – often referred to as the Dunning-Kruger effect – has far reaching implications because, presumably, one of the prerequisites of voluntary self-improvement is actually recognizing the need for improvement. Indeed, those who would benefit the most from understanding the limits of their reasoning are the ones who are least likely to do so. From a theoretical standpoint, estimating the degree of miscalibration among the incompetent is also paramount to understanding the causes and consequences of biases in judgment and decision making writ large. In the present work, we examine incompetence in the realm of biased or intuitive responding in reasoning tasks and provide evidence that Dunning-Kruger effects extend to self-reported analytic thinking disposition.

Dual-process theory and analytic thinking disposition.

The distinction between autonomous/ intuitive (Type 1) processes and deliberative/ analytic (Type 2) processes, as formalized in dual-process theories (De Neys, 2012; Evans & Stanovich, 2013; Handley & Trippas, 2015; Kahneman, 2011; Pennycook, Fugelsang, & Koehler, 2015b; Sloman, 2014; Thompson, Prowse Turner, & Pennycook, 2011), is now exceedingly popular in many areas of psychology. One of the key implications of this distinction is that analytic thinking proceeds via at least some form of volitional control, which in turn indicates that people will differ in terms of their disposition toward analytic thinking. Put simply, some people appear to be more willing to engage in deliberative thought than others (Stanovich & West, 1998; Stanovich & West, 2000; Stanovich, 2012). Recent research has shown that

individual differences in analytic thinking disposition (also referred to as "analytic cognitive style") are consequential for a wide range of psychological domains over and above individual differences in cognitive ability or intelligence (Pennycook, Fugelsang, & Koehler, 2015a).

Analytic thinking disposition is also assessed using self-report scales, such as the Need for Cognition (NC) scale (Cacioppo & Petty, 1982; Cacioppo, Petty, Feinstein, & Jarvis, 1996; Epstein, Pacini, Denes-Raj, & Heier, 1996; Fleischhauer et al., 2010). According to Petty, Brinol, Loersch, and McCaslin (2009), Need for Cognition “refers to the tendency for people to vary in the extent to which they engage in and enjoy effortful cognitive activities” (p. 318). It includes items such as “I am not very good at solving problems that require careful logical analysis” and “I enjoy solving problems that require hard thinking”. Thus, “effortful cognitive activities” in Petty et al.’s (2009) definition refers specifically to thinking and problem solving.

Pacini and Epstein (1999) further specified the NC scale by creating subscales that distinguish between ‘ability’ and ‘engagement’. This distinction can be seen in the two examples offered above: NC includes items that index both the self-reported ability to engage in effortful thought and the enjoyment derived therefrom. Contrast, for example, the item “I am much better at figuring out things logically than other people” (ability subscale) with the item “I try to avoid situations that require thinking in depth about something” (engagement subscale). Thus, Pacini and Epstein’s NC scale distinguishes two important components of general analytic thinking dispositions in a way that parallels performance-based measures.

NC has been used frequently in studies on information evaluation and recall, attitude formation, and judgment and decision making (see, Petty et al., 2009, for review), along with a variety of other domains not traditionally associated with information processing. For example, high NC is associated with decreased religious and paranormal belief (Pennycook, Cheyne, Barr,

Koehler, & Fugelsang, 2014; Svedholm & Lindeman, 2013), increased life satisfaction (Gauthier, Christopher, Walter, Mourad, & Marek, 2006), decreased support for punitive responses to crime (Sargent, 2004), and utilitarian moral judgment (Conway & Gawronski, 2013). Indeed, Petty et al. (2009) note that over 1,000 publications have either cited the article that introduced the NC scale (Cacioppo & Petty, 1982) or the article that introduced a shortened version of the scale (Cacioppo, Petty, & Kao, 1984).

Methodological implications of the Dunning-Kruger effect.

Although the NC scale has been used in hundreds of studies (Cacioppo et al., 1996; Petty et al., 2009), research on self-report thinking disposition faces a dilemma: People who are genuinely unwilling to engage analytic thinking may not be well suited to estimate their degree of NC. Indeed, there is evidence that high NC is associated with increased levels of thinking about one's own thinking (Petty et al., 2009). Moreover, Kruger and Dunning (1999) found that people who were in the bottom quartile in terms of logical reasoning estimated that they were just above average in terms of accuracy whereas, if anything, those in the top quartile somewhat underestimated their performance. Mata, Ferreira, and Sherman, (2013) demonstrated that relatively analytic people have a metacognitive advantage over those who rely primarily on their intuition because they are aware of *both* the intuitive answer and the deliberative (typically correct) response. Similar to Kruger and Dunning (1999), Mata et al. (2013) found that intuitive reasoners strongly overestimate their performance relative to deliberative reasoners. Given that people's "chronic self-views" (i.e., opinion about one's abilities independent of actual performance) have been shown to influence their performance (Atir, Rosenzweig, & Dunning, 2015; Critcher & Dunning, 2009; Ehrlinger & Dunning, 2003), previous research suggests that

people overestimate their performance on high-level reasoning tasks precisely because they view themselves as more reasonable than is justified by their objective performance.

Based on this research, we hypothesize that self-report thinking disposition scales are miscalibrated in a systematic way, such that those who are genuinely not analytic should overstate their relative analyticity, whereas people who are genuinely analytic should fairly accurately report their NC or perhaps even underreport it relative to others in the sample. In other words, the association between NC and objective performance should be similar to the association between estimated and objective performance: Overestimation should be largest among the most biased and the smallest (or reversed) among the least biased.

Theoretical implications of the Dunning-Kruger effect.

The Dunning-Kruger effect has important – although heretofore unspecified – theoretical implications for recent work in the field of heuristics and biases in reasoning (but see, Mata, Ferreira, et al., 2013 for related empirical work). Namely, there is growing evidence that people can recognize (if implicitly) the conflict inherent in many reasoning problems (see De Neys, 2012; De Neys, 2014 for reviews). For example, consider the following item from the Cognitive Reflection Test (Frederick, 2005):

A bat and ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost?

Typically, around 65% of participants respond ‘10 cents’ to this problem (e.g., Pennycook, Cheyne, Koehler, & Fugelsang, 2016) even though it is incorrect (if the ball costs 10 cents, the bat must cost \$1.10 and in total they would cost \$1.20). This is thought to occur because people are cognitive misers (i.e., they conserve mental resources when possible, see Toplak, West, & Stanovich, 2011) and unduly rely on the first thing that comes to mind.

If the Dunning-Kruger effect applies to this type of problem, those who give incorrect intuitive responses would be unlikely to recognize their bias (Mata, Ferreira, et al., 2013). Surprisingly, however, there is some evidence that even people who get this problem incorrect have some sense that there is something “off” about the problem. Namely, participants who give the incorrect intuitive response to conflict problems are less confident relative to when they answer a non-conflict version of the task (i.e., one that does not cue an incorrect intuitive response; De Neys et al., 2013); a pattern of results that remains even after participants are put under cognitive load (Johnson, Tubau, & De Neys, 2016). Such conflict detection has been evidenced using a variety of different measures (e.g., response time, De Neys & Glumicic, 2008; memory recall, Franssens & De Neys, 2009; skin conductance response, De Neys, Moyens, & Vansteenwegen, 2010; fMRI, De Neys, Vartanian, & Goel, 2008; and ERP, Banks & Hope, 2014) and a wide range of tasks (e.g., syllogisms, De Neys & Franssens, 2009; the conjunction fallacy, De Neys, Cromheeke, & Osman, 2011; and ratio bias, Bonner & Newell, 2010).

Conflict detection during reasoning has been referred to as “omnipresent, regardless of whether participants answer problems correctly or incorrectly” (De Neys et al., 2008, p. 488). Indeed, there is evidence for conflict detection even among particularly biased participants using subtle low-level measures such as skin conductance (De Neys et al., 2010). Interestingly, however, this line of research does not appear consistent with the strong evidence for the Dunning-Kruger effect. If people are good at detecting conflict during reasoning, why are the incompetent so unaware of their incompetence?

One possibility is that the low-level detection of conflicting outputs, however efficient, may not translate into changes in behavior and, ultimately, reductions in biased responding. Indeed, recent research indicates that failures of conflict detection (which may be due to either a

lack of conflict detection signal or a lack of responsiveness to a present conflict signal) are more common than previously thought (Pennycook, Fugelsang, et al., 2015b, 2012). Moreover, there is evidence that less analytic individuals are less likely to respond to conflict during reasoning (Mevel et al., 2015; Pennycook et al., 2014, 2015b; Thompson & Johnson, 2014). Thus, metacognitive monitoring may be more effective among genuinely analytic individuals regardless of the effectiveness of conflict detection per se. This suggests that non-analytic participants should overestimate their accuracy on problems that include an intuitive, yet incorrect lure, like the bat and ball example above (Mata, Ferreira, et al., 2013). More generally, the foregoing indicates that analytic individuals may be better suited to assess their relative degree of analyticity.

The current work

We report two studies in which participants completed a popular performance-based measure of analytic thinking disposition, the Cognitive Reflection Test (CRT; Frederick, 2005), and were subsequently asked to estimate how many of the items they got correct (Mata, Ferreira, et al., 2013; Noori, 2016). Following Kruger and Dunning (1999), we hypothesized that participants who perform poorly on the CRT will overestimate their performance to a greater extent than those who perform well (i.e., less analytic people should be more poorly calibrated). In addition, participants were also asked to self-report their need or desire to think analytically using the NC scale. We predicted a Dunning-Kruger effect such that participants who perform particularly poorly on the CRT (indicating an intuitive or non-analytic thinking disposition) would over-report the degree to which they are disposed to analytic thinking. In Study 2, we used an independent assessment of analytic thinking – the heuristics and biases inventory (Toplak, West, & Stanovich, 2014; Toplak et al., 2011) – to assess whether non-analytic individuals are

genuinely worse at recognizing their bias. Put differently, participants who are decidedly non-analytic based on the performance measure should be less suited to assess their degree of analyticity on a self-report measure, leading to poor calibration both in terms of estimated CRT accuracy and self-reported NC.

Study 1

Kruger and Dunning's (1999) primary finding was a larger difference between actual performance and estimated performance (i.e., miscalibration in the form of overestimation) for the incompetent relative to the competent. Thus, participants who perform relatively poor on the CRT should overestimate their performance and those who do relatively well should be better calibrated (Mata, Ferreira, et al., 2013), which suggests that those who are genuinely non-analytic may not be aware of their lack of analyticity.

Assuming there is good evidence for overconfidence on the estimated CRT task, we can then assess the association between this miscalibration and self-reported analytic thinking style (via the NC scale). If the tendency to overestimate CRT performance translates into a tendency to rate oneself relatively high in Need for Cognition, estimated CRT performance should predict NC over and above actual CRT performance. Similarly, the well-established association between CRT and NC (e.g., Pennycook, Cheyne, Koehler, & Fugelsang, 2015) should not be evident once CRT scores have been calibrated by subtracting estimated CRT scores. Furthermore, these findings should hold for both ability and engagement subscales of the NC scale.

Method

In all studies, we report how we determined our sample size, all data exclusions, and all measures. Data are available online via the Open Science Framework: <https://osf.io/syfn8/>.

Participants

We recruited our participants through Amazon's Mechanical Turk and chose a target N of 200 to have sufficient power (90%) to detect a moderate effect size ($r = .20$). We removed 14 participants because they responded affirmatively when asked if they had responded randomly at any point during the survey. We also removed a participant who had missing values for the CRT questions. Two additional participants were removed because they gave numerical answers to a CRT question that required a non-numerical response. The resulting sample ($N = 183$, *Mean* age = 33.7, *SD* age = 9.5) consisted of 99 males and 84 females.

Materials and Procedure

Cognitive Reflection Test. Frederick's (2005) original CRT consisted of 3 math problems that reliably cue intuitive but incorrect answers. The test is considered an index of analytic thinking disposition because the items require the participant to question a compelling intuitive response – a process that requires a willingness to think analytically (Pennycook & Ross, 2016; Travers, Rolison, & Feeney, 2016). Since the original CRT has been used extensively on Mechanical Turk, we used two recently developed tests that were designed to measure the same underlying construct instead of the original measure. Specifically, we used Toplak, West, and Stanovich's (2014) 4-item CRT and Thomson and Oppenheimer's (2016) 4-item CRT. The number of correct responses for the two CRTs were significantly correlated, $r(183) = .336$, $p < .001$, and combined to make a scale with acceptable reliability, $\alpha = .70$. The eight CRT items were presented one at time and in a random order for each participant. After completing all CRT items, we asked participants to estimate the number of CRT questions that they answered correctly.

Need for Cognition. Participants then completed Pacini and Epstein's (1999) 20-item Need for Cognition scale. Pacini and Epstein's NC scale consists of 'ability' and 'engagement'

subscales, which were treated as separate scales for present purposes. Items were presented in a randomized order and participants responded on a scale from “1) Definitely not true of myself” to “5) Definitely true of myself”. Both Ability (NC-A) and Engagement (NC-E) subscales had excellent reliability, α 's = .90, .95 (for NC-A and NC-E, respectively). The full NC scale was also reliable, $\alpha = .95$

Results and Discussion

Dunning-Kruger effects in estimated CRT performance

As predicted, participants overestimated their total accuracy on the eight CRT items. On average, participants estimated that they correctly solved 5.59 CRT problems ($SD = 1.52$), but mean performance was only 3.88 ($SD = 2.11$), $t(182) = 11.14$, $SE = .15$, $p < .001$, $d = .82$. The correlation between estimated and actual CRT performance was modest, $r(183) = -.379$, $p < .001$, such that *actual* CRT performance only explains 14.4% of the variance in *estimated* CRT performance. Following Kruger and Dunning (1999), we split the sample into groups, based on accuracy, to test whether those who did poorly on the CRT were more strongly miscalibrated than those who did better. Although there are only 9 possible scores on the 8-item CRT and, in theory, we could use 9 different groups, only a small number of participants scored 0 ($N = 4$, 2.2% of the sample) or 8 ($N = 8$, 4.4% of the sample). Thus, we increased the N in each CRT group by combining across accuracy scores and creating four groups (0-2, 3-4, 5-6, 7-8).

Using a mixed design ANOVA, there was an interaction between CRT group and the difference between actual CRT score and estimated CRT score (see Figure 1), $F(3, 179) = 56.24$, $MSE = 1.13$, $p < .001$, $\eta^2 = .49$. As is evident from Figure 1, overestimation decreases (i.e., calibration increases) systematically as accuracy increases. To compare rates of calibration across CRT categories, we computed a difference score between estimated and actual CRT and

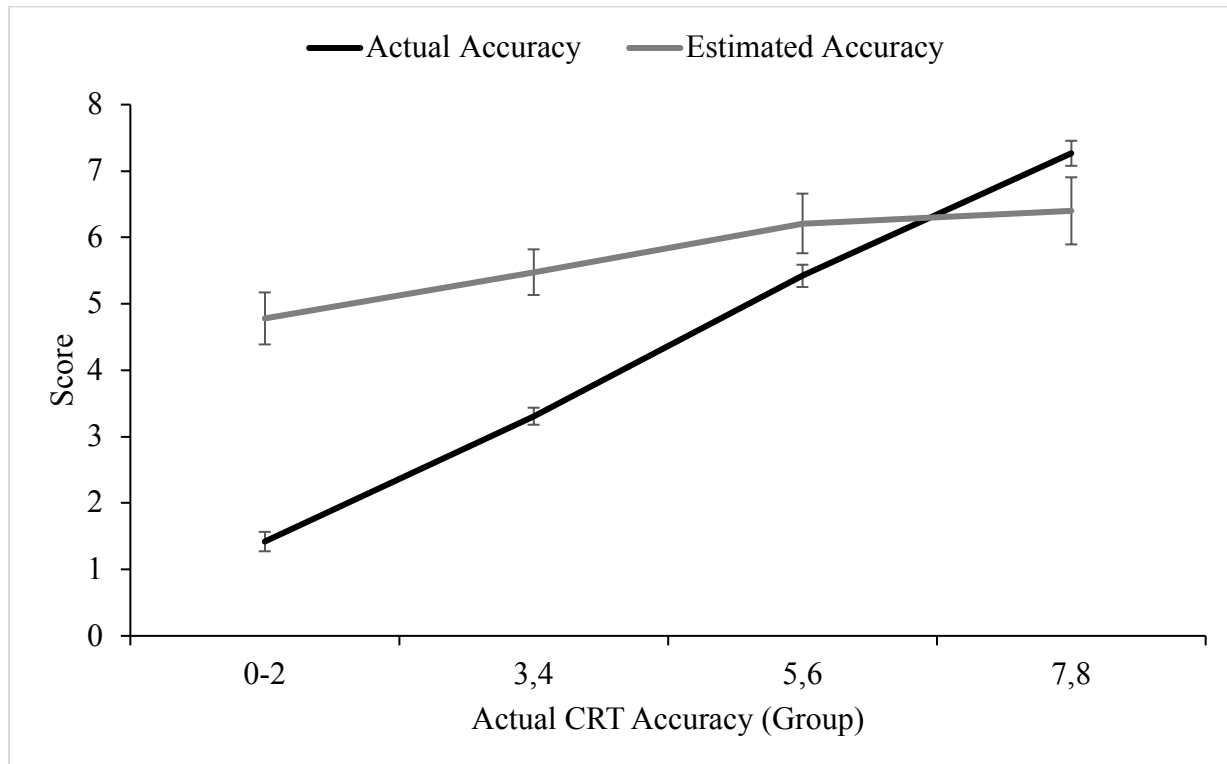
compared it across the levels of CRT performance (based on our 4 groups). A *post hoc* Tukey HSD test comparing the difference between estimated and actual CRT performance (i.e., calibration) indicated that each of the four groups emerged as separate homogeneous subsets ($p < .05$). Notably, those who scored very low on the CRT (0-2, $M = 1.42$, 17.8% accuracy) estimated that they answered 4.78 questions correctly (59.8% accuracy) – that is, they overestimated by a factor of 3.4. Those who correctly answered 3-4 of the CRT problems overestimated by a factor of 1.7, whereas those who answered 5-6 problems correctly only overestimated by a factor of 1.1 (see Figure 1).

In contrast, and similar to Kruger and Dunning (1999), we found that those who scored 7-8 *underestimated* their performance by a factor of 1.1, $t(29) = 3.21$, $p = .003$, $d = .59$. This was not driven by the few participants ($N = 8$) who scored a perfect 8 out of 8 (and for whom it would be impossible to overestimate performance – those who scored 7 out of 8 ($N = 22$) estimated that they got, on average, 6.36 ($SD = 1$) out of 8 correct (which was significantly lower than their true score of 7, $t(21) = 2.98$, $p = .007$, $d = .64$). Importantly, those who scored 5-6 out of 8 significantly *overestimated* their performance, $t(37) = 4.35$, $p < .001$, $d = .71$, indicating that the monotonic improvement in calibration as CRT performance increases cannot be attributable to those who score too high for overestimation to be possible. Moreover, it should be noted that these results cannot be attributed to mere regression to the mean. Particularly intuitive individuals (those who scored 0-2 out of 8) overestimated by a factor of 3.4 whereas particularly analytic individuals (those who scored 7-8 out of 8) underestimated by a factor of 1.1. That is, the reported misalignment is asymmetric in a way predicted by previous research.

Figure 1

Misalignment between actual CRT accuracy and estimated CRT accuracy as a function of CRT group (those who score 0-2, 3-4, 5-6, or 7-8 out of 8) in Study 1. Y-Axis represents the number

of correct/estimated correct out of 8 CRT problems. Error bars represent 95% confidence intervals. Group 1 (0-2) $N = 50$; Group 2 (3-4) $N = 65$; Group 3 (5-6) $N = 38$; Group 4 (7-8) $N = 30$.



Association between overconfidence and Need for Cognition

If the NC scale is a good measure of actual analytic thinking disposition, it should correlate more strongly with actual CRT performance than estimated performance. However, if anything, *estimated* CRT performance was more strongly correlated with NC, $r(183) = .307, p < .001$, than was *actual* CRT performance, $r(183) = .246, p = .001$. This was more evident for the ability subscale (NC-A), which correlated with estimated CRT performance nearly as strongly as actual CRT performance, $r(183) = .352, p < .001$, but was more modestly correlated with actual CRT performance, $r(183) = .242, p = .001$. However, there was no difference between the magnitude of these correlations via a William's test, $t(180) = 1.42, p = .159$. The correlation between the engagement subscale (NC-E) and estimated CRT performance, $r(183) = .229, p =$

.002, was more similar to actual CRT performance, $r(183) = .216, p = .003$. Thus, if anything, miscalibration was strongest when participants self-reported their ability to engage analytic thinking.

If the tendency to overestimate CRT performance is linked with the tendency to rate oneself relatively high on the NC scale, estimated CRT should predict NC over and above actual CRT performance. We therefore entered both accuracy and estimated accuracy (along with their interaction) as predictors in two regression analyses with NC-A and NC-E as the dependent variables (see Table 1). Not only did estimated CRT accuracy significantly predict self-reported NC-A and NC-E once actual CRT scores were taken into account, but actual CRT accuracy was not a robust predictor once *estimated* accuracy was taken into account. There was also an interaction between estimated accuracy and actual accuracy for both NC subscales. This interaction emerged because estimated CRT accuracy was more strongly positively correlated with NC for those who did well on the CRT (see Figure S1 in supplementary materials). These results indicate a link between miscalibration in analytic thinking and self-reported Need for Cognition. As a follow-up analysis, we created a calibration score by taking the difference between estimated and actual CRT. This calibration score was not significantly correlated with NC-A, $r(183) = -.012, p = .871$, or NC-E, $r(183) = .051, p = .491$. This, again, indicates that self-report NC is as much a measure of estimated CRT performance as it is a measure of actual CRT performance.

Table 1

Final step of hierarchical multiple regression analysis predicting self-reported Need for Cognition with estimated CRT accuracy (CRT Est), actual CRT accuracy (CRT Acc), and their interaction (Acc x Est) as predictors. The top panel is the ability subscale (NC-A) and the bottom panel is the engagement subscale (NC-E). $N = 183$.

(NC-A)	<i>r</i>	B	<i>t</i>	<i>P</i>	<i>Tolerance</i>	<i>VIF</i>
Intercept			0.92	.358		

CRT Acc	.242	.101	1.35	.178	.84	1.19
CRT Est	.352	.326	4.40	< .001	.85	1.18
Acc x Est	.166	.179	2.59	.010	.98	1.02
(NC-E)	<i>r</i>	B	<i>t</i>	<i>P</i>	<i>Tolerance</i>	<i>VIF</i>
Intercept			0.86	.392		
CRT Acc	.216	.126	1.63	.106	.84	1.19
CRT Est	.229	.193	2.50	.013	.85	1.18
Acc x Est	.169	.170	2.38	.019	.98	1.02

Study 2

In Study 1, those who were prone to errors on the CRT were more likely to overestimate their performance – a Dunning-Kruger effect. Moreover, consistent with previous research in other domains (Atir et al., 2015; Critcher & Dunning, 2009; Ehrlinger & Dunning, 2003), estimates of CRT accuracy were as predictive of self-reported analytic thinking disposition as was actual CRT accuracy. These findings suggest that self-report measures of thinking disposition may lack precision. Overconfidence among those genuinely low in analytic thinking may translate into overestimates of self-report analytic thinking whereas proper calibration (or, if anything, under-confidence) among those genuinely high in analytic thinking may translate into relative underestimates of self-report analytic thinking. In other words, it may be the case that those who rely on their intuition are not analytic enough to know that they are not analytic whereas analytic people are analytic enough to know the limits of their analyticity.

Although the results of Study 1 were consistent with the proposed Dunning-Kruger effect in self-reported thinking disposition, the evidence was indirect. Specifically, the same performance-based analytic thinking disposition measure, the CRT, was used to make inferences about miscalibration for both estimated CRT score and self-reported NC. To overcome this limitation, we included a separate measure of analytic thinking: the heuristics and biases inventory (H&B; Toplak, West, & Stanovich, 2011). Having a second performance-based

(‘objective’) measure of analytic thinking will allow us to more directly test the hypothesis that NC scores are systematically miscalibrated. That is, the difference between actual and estimated CRT performance (calibration) should be associated with objective H&B performance, but not self-reported NC. This would illustrate a correspondence between objective measures of analytic thinking even after miscalibration has been taken into account. There should be no likewise correspondence with self-report NC because, presumably, this miscalibration also affects people’s perception of their analytic thinking disposition.

Including a second objective measure of analytic thinking will also allow us to compare *relative* performance on our two analytic thinking benchmarks (CRT and H&B) with *relative* self-reported Need for Cognition by creating a number of additional calibration scores. Individuals who are not particularly analytic (based on relative CRT and H&B score) should nonetheless rate themselves as relatively analytic (based on self-reported NC). We predict that the difference between relative CRT performance and self-reported NC will correlate positively with H&B performance (and, likewise, the difference between H&B and NC will correlate with CRT). In other words, there should be miscalibration between NC and our objective measures of analytic thinking in the same way that there is miscalibration between estimated and actual CRT scores. This would be akin to a Dunning-Kruger effect in self-reported analytic thinking disposition. As in Study 1, this should be evident for both ability and engagement subscales of the NC scale.

Method

Data are available online via the Open Science Framework: <https://osf.io/z6trc/>.

Participants

In total, 400 participants were recruited from Mechanical Turk¹. We removed 56 participants because they responded affirmatively when asked if they had responded randomly at any point during the survey and 3 participants who had missing data for the CRT. The resulting sample ($N = 341$, $Mean\ age = 33.9$, $SD\ age = 10.5$) consisted of 200 males and 139 females.

Materials and Procedure

Cognitive Reflection Test. The CRT was administered as in Study 1. The two CRT scales were significantly correlated, $r(341) = .362$, $p < .001$, but the scale had weaker reliability than in Study 1, $\alpha = .65$. One participant estimated that they solved 9 out of 8 CRT problems correctly, so we changed the value to 8.

Heuristics and biases battery. As an independent measure of analytic thinking, we used Toplak et al.'s (2011) heuristics and biases battery. The scale consists of 16 problems that were derived from Kahneman and Tversky's heuristics and biases research program (see Kahneman, 2011 for a review). Items include biases such as the gambler's fallacy and the conjunction fallacy (see appendix in Toplak et al., 2011). The scale had weak reliability, $\alpha = .68$.

Need for Cognition. The NC scale was administered as in Study 1, except in this case it followed the heuristics and biases battery and was presented before a paranormal belief scale (which will not be considered further, see supplementary materials). One participant had an outlying NC score (3 SD below the mean) – subsequent analyses are reported with the outlier

¹ We originally collected data for 200 participants using the same stopping rule as in Study 1. However, upon completing data collection for the first 200 participants in this study, an error in the CRT estimate question was discovered (which also applied to Study 1). Specifically, participants were told to input a number between 1 and 8 instead of between 0 and 8. We therefore re-ran Study 2 with the corrected instructions. Fortunately, only one participant guessed that they got 0 out of 8 correct when given the corrected instructions (no participants indicated 0 in Study 1 or the first run of Study 2). Estimated CRT performance was identical regardless of whether participants were told to enter a number from 1 to 8 or 0 to 8, $t(339) = .04$, $p = .971$. Thus, we report the data for the full set of 400 participants below.

removed, although the pattern of results was identical. The subscales were reliable, $\alpha = .89, .94$ for NC-A and NC-E, respectively.

Results and Discussion

Dunning-Kruger effects in estimated CRT performance

Estimated CRT accuracy ($M = 5.62, SD = 1.59$) was greater than actual accuracy ($M = 3.85, SD = 1.98$), $t(340) = 16.42, SE = .11, p < .001, d = .89$. As in Study 1, we split the sample into four groups based on accuracy of CRT scores to examine the relationship between actual and estimated CRT scores. Using a mixed design ANOVA, there was an interaction between low/high CRT group and the difference between actual CRT score and estimated CRT score (see supplementary Figure S2), $F(3, 337) = 92.64, MSE = 1.09, p < .001, \eta^2 = .45$. A *post hoc* Tukey HSD test comparing the difference between estimated and actual CRT performance (i.e., calibration) indicated that all four groups emerged as separate homogeneous subsets ($p < .05$). This indicates a decrease in overestimation at each increasing level of CRT performance. As in Study 1, those who scored lowest on the CRT (0-2 correct) overestimated their performance by a factor of 3.4. Even those who scored 5-6 out of 8 significantly overestimated their performance, $t(52) = 6.63, p < .001, d = .67$. In contrast, those who scored highest (7-8 correct) significantly *underestimated* their performance, $t(33) = 3.86, p = .001, d = .66$.

Association between overconfidence and Need for Cognition

Here we will replicate the associations between estimated and actual CRT performance and NC. Correlations among primary variables can be found in Table 2. As in Study 1, the correlation between NC and actual CRT performance were not significantly different (although in this case, nominally larger for NC-E) to the correlation between NC and estimated CRT

performance, $t < 1^2$. To estimate the extent to which actual CRT accuracy and estimated CRT accuracy predict NC, we entered both variables (along with their interaction) as predictors in two regression analyses with NC-A and NC-E as dependent variables (see Table 3). Unlike in Study 1, both actual and estimated CRT accuracy (along with their interaction, see Figure S3 in supplementary materials) significantly independently predicted NC-A and only actual CRT significantly predicted NC-E. These independent associations were relatively modest, however; β 's ranged from .120-.181.

² Supernatural (religious and paranormal) belief was more strongly correlated with actual CRT accuracy, $r(341) = -.278, p < .001$, than estimated CRT accuracy, $r(341) = -.093, p = .086, t(338) = 3.23, p = .001$. This provides an existence proof for the idea that actual CRT can be more predictive than estimated CRT accuracy.

Table 2. Correlations (Pearson r) among primary variables in Study 2. $N = 341$.

	1	2	3	4	5	6	7	8	9
1. CRT Accuracy	-								
2. Estimated CRT Accuracy	.400***	-							
3. Need for Cognition - Ability	.202***	.203***	-						
4. Need for Cognition - Engagement	.190***	.133*	.672***	-					
5. Heuristics & Biases	.411***	.326***	.194***	.203***	-				
6. CRT ₁ – Estimated Accuracy ₂ (Calibration1)	.676***	-.404***	.038	.082	.148**	-			
7. CRT ₁ – NC-A ₃ (Calibration2)	.632***	.156**	-.632***	-.382***	.172**	.505***	-		
8. H&B ₄ – NC-A ₃ (Calibration3)	.165**	.097	-.635***	-.369***	.635***	.086	.633***	-	
9. CRT ₁ – H&B ₄	.543***	-.068	-.007	.012	.543***	-.487***	-.424***	.433***	-

***indicates $p < .001$, **indicates $p < .01$, *indicates $p < .05$

Table 3

Final step of a hierarchical multiple regression analysis predicting self-reported Need for Cognition with estimated CRT accuracy (CRT Est), actual CRT accuracy (CRT Acc), and their interaction (Acc x Est) as predictors. The top panel is the ability subscale (NC-A) and the bottom panel is the engagement subscale (NC-E). $N = 341$.

(NC-A)	r	B	t	p	Tolerance	VIF
Intercept			0.86	.389		
CRT Acc	.202	.122	2.11	.036	.82	1.23
CRT Est	.203	.181	3.04	.003	.78	1.29
Acc x Est	.087	.120	2.20	.028	.92	1.08
(NC-E)	r	B	T	p	Tolerance	VIF
Intercept			0.44	.657		
CRT Acc	.190	.151	2.56	.011	.82	1.23
CRT Est	.133	.087	1.44	.150	.78	1.29
Acc x Est	.052	.063	1.13	.258	.92	1.08

As in Study 1, neither NC-A or NC-E were associated with the difference between actual and estimated CRT performance, r 's < .085, p 's > .130 (Calibration1; see Table 2). In contrast, H&B performance *did* positively associate with Calibration1, $r(341) = .148, p = .006$. Thus, there was a correspondence between objective measures of analytic thinking (CRT and H&B) even after miscalibration (CRT estimates) have been taken into account. This pattern of results was not evident for self-reported NC because, presumably, the miscalibration that is reflected in estimated CRT performance also affects people's perception of their analytic thinking disposition.

Dunning-Kruger effects in self-reported Need for Cognition

To further test the hypothesis that participants who are genuinely low in analytic thinking over-report their NC, we created three additional calibration scores. We did this by converting CRT, H&B, and NC raw scores into z-scores and computed their difference scores (Table 2). The conversion to z-scores allows us to compare participants' *relative* position on the three ostensibly linked measures. The goal is to treat self-reported NC in the same way as estimated

CRT scores. For this analysis, we decreased the number of comparisons by focusing on the ability subscale (NC-A), which more directly relates to people's assessments of their analytic thinking disposition (as opposed to mere enjoyment of analytic thinking)³.

If NC-A is miscalibrated in the same way as estimated CRT scores, as we have predicted, there should be a correspondence between the parallel estimate-based and NC-based calibration scores. Indeed, the difference between the z-scores for CRT accuracy and NC-A correlated positively with H&B performance ('Calibration2'; Table 2), $r(341) = .172, p = .001$.

Correspondingly, the difference between the z-scores for H&B accuracy and NC-A correlated positively with CRT performance ('Calibration3'; Table 2), $r(341) = .165, p = .002$. Participants who were less analytic based on H&B performance were more likely to rate themselves as *relatively* higher in NC than would be warranted given their CRT performance (and vice versa). There was no corresponding correlation between NC-A and the difference between the z-scores for CRT and H&B, $r(341) = -.007, p = .894$, which indicates a dissociation between self-reported and performance-based measures of analytic thinking.

This analysis illustrates that self-reported Need for Cognition (ability subscale) is more similar (in terms of its correlates) to *estimated* Cognitive Reflection Test performance than to *actual* Cognitive Reflection Test *or* heuristics and biases performance. To summarize, self-reported Need for Cognition does *not* predict the difference between actual and estimated Cognitive Reflection Test performance, but heuristics and biases performance *does* predict this difference. Moreover, when the difference between relative (actual) Cognitive Reflection Test

³ The results are nonetheless essentially identical when NC-E is used in place of NC-A. Specifically, the difference between the z-scores for CRT accuracy and NC-E correlated positively with H&B performance (akin to 'Calibration2' in Table 2), $r(341) = .176, p = .001$. The difference between the z-scores for H&B accuracy and NC-E correlated positively with CRT performance (akin to 'Calibration3' in Table 2), $r(341) = .163, p = .002$

and Need for Cognition scores is used as a calibration score, heuristics and biases performance predicts that as well. Similarly, when the difference between heuristics and biases performance and Need for Cognition is used like a calibration score, Cognitive Reflection Test performance predicts *that*. Finally, to complete the full set of possible analyses, the difference between Cognitive Reflection Test and heuristics and biases performance is not predicted by Need for Cognition since it is not a meaningful difference (unlike the difference between CRT/H&B and NC scores, as hypothesized).

To illustrate the source of the Dunning-Kruger effect on self-reported NC, we will focus on the parallel between two independent patterns of results: 1) The difference between actual and estimated CRT scores as a function of CRT group (Figure 2): interaction, $F(3, 337) = 92.64$, $MSE = 1.09$, $p < .001$, $\eta^2 = .45$; and 2) The difference between H&B performance and self-reported NC as a function of CRT group (Figure 3): interaction, NC (full scale); $F(3, 337) = 2.92$, $MSE = .77$ $p = .034$, $\eta^2 = .03$; NC-A; $F(3, 337) = 2.79$, $MSE = .79$ $p = .040$, $\eta^2 = .02$; NC-E; $F(3, 337) = 2.87$, $MSE = .78$ $p = .012$, $\eta^2 = .03$. In both cases, there is an interaction such that estimated accuracy/self-report NC is larger than objective performance for those who score relatively low on the CRT, but this difference decreases and eventually reverses as CRT performance increases. Indeed, the comparison between H&B performance and NC indicates that, if anything, particularly analytic individuals are underreporting their NC relative to the remainder of the sample. It should be noted, however, that the confidence intervals overlap substantially at every level in this analysis despite the overall evidence for an interaction (see Figure 3).

Figure 2

Miscalibration between actual CRT accuracy and estimated CRT accuracy as a function of CRT group (those who score 0-2, 3-4, 5-6, or 7-8 out of 8) in Study 2. Y-Axis represents the number of correct/estimated correct out of 8 CRT problems. Error bars represent 95% confidence intervals. Group 1 (0-2) $N = 96$; Group 2 (3-4) $N = 113$; Group 3 (5-6) $N = 98$; Group 4 (7-8) $N = 34$.

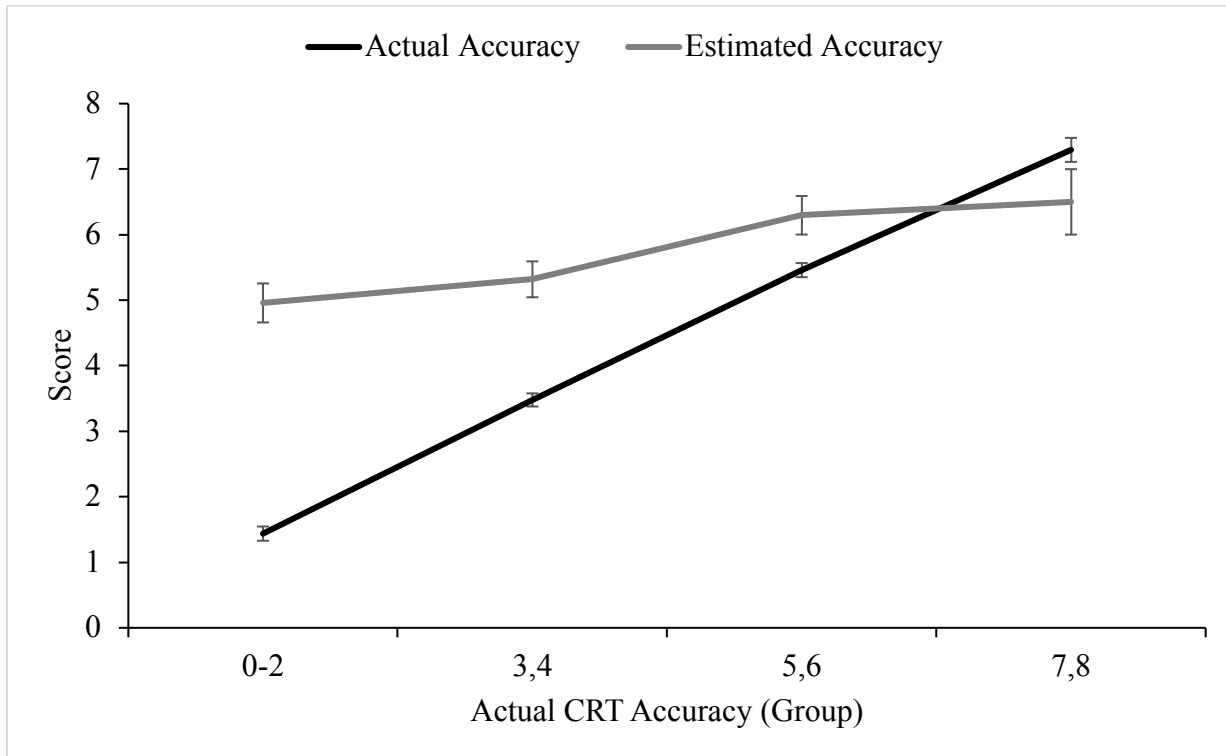
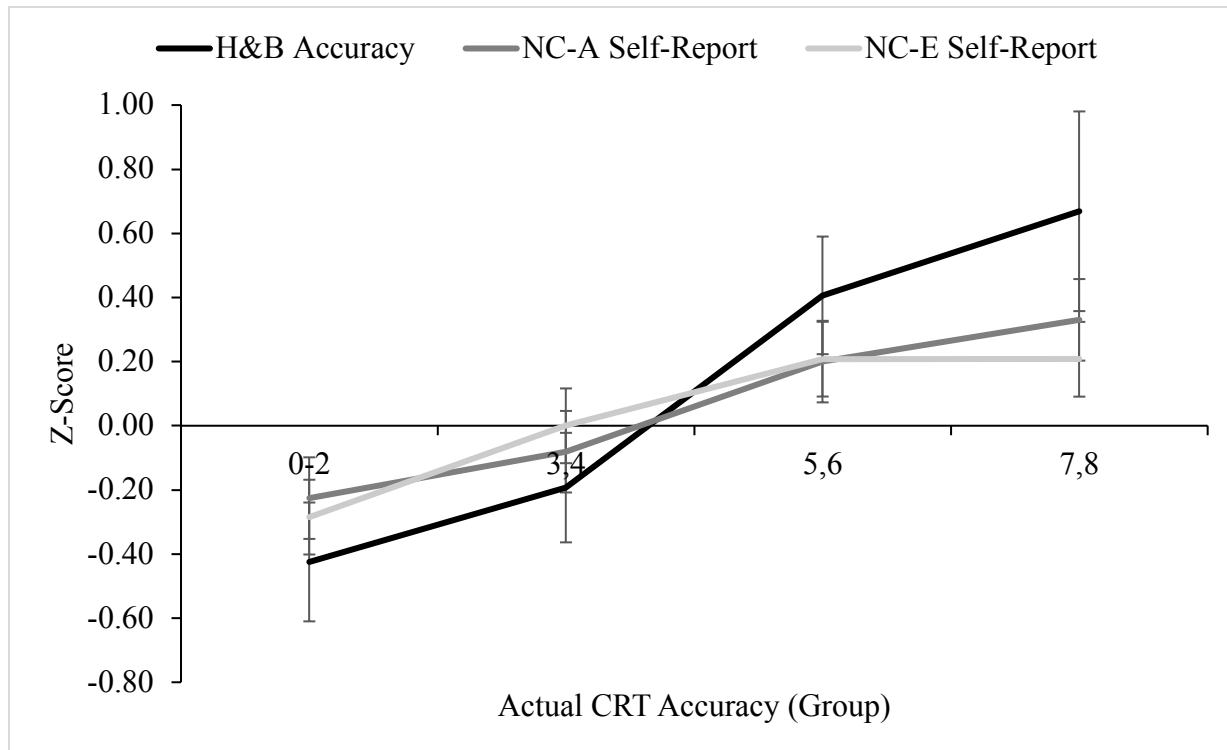


Figure 3

Miscalibration between heuristics and biases performance (H&B) and self-reported Need for Cognition (ability, NC-A; engagement, NC-E) as a function of CRT group (those who score 0-2, 3-4, 5-6, or 7-8 out of 8) in Study 2. Y-Axis represents z-scores. Error bars represent 95% confidence intervals. Group 1 (0-2) $N = 96$; Group 2 (3-4) $N = 113$; Group 3 (5-6) $N = 98$; Group 4 (7-8) $N = 34$.



General Discussion

Our results provide empirical support for Dunning-Kruger effects in both estimates of reasoning performance and self-reported thinking disposition. Particularly intuitive individuals greatly overestimated their performance on the Cognitive Reflection Test (CRT) – a tendency that diminished and eventually reversed among analytic individuals. Moreover, self-reported analytic thinking disposition – as measured by the ‘ability’ and ‘engagement’ subscales of the Need for Cognition (NC) scale – was just as strongly (if not more strongly) correlated with estimated CRT performance than actual CRT performance. In addition, an analysis using an

additional performance-based measure of analytic thinking – the heuristics and biases battery (H&B) – revealed a systematic miscalibration on self-reported NC wherein relatively intuitive individuals report that they are more analytic than is justified by their objective performance. Together, these findings indicate that participants who are low in analytic thinking (so-called “intuitive thinkers”) are at least somewhat unaware of (or unresponsive to) their propensity to rely on intuition in lieu of analytic thought during decision making. This conclusion is consistent with previous research that suggests that the propensity to think analytically facilitates metacognitive monitoring during reasoning (Pennycook, Fugelsang, & Koehler, 2015b; Thompson & Johnson, 2014). Those who are genuinely analytic are aware of the strengths and weaknesses of their reasoning, whereas those who are genuinely non-analytic are perhaps best described as “happy fools” (De Neys et al., 2013).

This research has both methodological and theoretical implications. With respect to methodological implications, self-report measures of thinking disposition such as the NC scale have been used in hundreds of studies (Cacioppo et al., 1996; Petty et al., 2009). Nonetheless, correlations with NC are often modest. In the current work, for example, correlations (Pearson’s r) between performance-based measures of analytic thinking and the ‘ability’ and ‘engagement’ NC subscales ranged from .190 to .242. Given the evidence for systematic miscalibration among genuinely non-analytic individuals, we suggest that the predictive power of analytic thinking dispositions may have been underestimated in research that has focused on self-report measures. Future research should continue to explore whether performance-based measures of analytic thinking yield stronger associations than self-report measures.

In terms of theoretical implications, there has been a recent surge in metacognitive perspectives in the realm of dual-process theory. For example, Thompson and colleagues

(Thompson et al., 2011, 2013) have provided evidence that ‘feelings of rightness’ are predictive of the extent and quality of analytic thinking. Mata, et al. (2013) found a metacognitive advantage for deliberative reasoners because they are aware of both the intuitive and reflective responses to reasoning problems. In contrast, much research indicates that people are capable of detecting conflict between competing reasoning outputs (see De Neys, 2012; De Neys, 2014 for reviews). Indeed, as mentioned earlier, conflict detection has been referred to as “omnipresent” (De Neys et al., 2008, p. 488) and even particularly biased participants have been shown to have an increased skin conductance response when faced with a conflict-inducing reasoning problem (De Neys et al., 2010). This indicates that even the most biased of reasoners can be sensitive to stimuli that cue conflicting responses.

The contrast between these lines of research is particularly stark when it comes to the Cognitive Reflection Test. Whereas De Neys et al. (2013) found that even people who gave the intuitive response to the bat and ball problem (as highlighted in the introduction) were only around 82% confident in their response (compared to 97% confidence on a control version of the problem), the present results indicate that those who perform poorly on the CRT massively overestimate their accuracy (i.e., by a factor > 3). This contraposition indicates that the neurological (e.g., De Neys et al., 2008) and physiological (De Neys et al., 2010) conflict detection signal may be relatively effective, but the response to this signal may actually be rather ineffective. De Neys et al. (2013) found a large (15%) decrease in confidence for the bat and ball problem relative to a control, but 82% confidence is still quite high.

There is also evidence from verbal protocols that conflict detection is often implicit (De Neys & Glumicic, 2008) and it seems likely that it may only become explicit upon further reflection via analytic processing. Thus, given that analytic thinking relies (to some extent) on

volitional control, the presence of a signal to think analytically does not guarantee that the individual will engage in more than cursory levels of analytic thought. This line of reasoning is supported by both the present findings and previous work showing that the propensity to think analytically correlates with increases in response time for biased responses to incongruent (conflict) base-rate problems (Pennycook et al., 2014; Pennycook, Fugelsang, & Koehler, 2015b). That is, more analytic individuals engaged in more substantive analytic thinking even in cases where they ultimately rationalized their initial ('biased') response. Collectively, these findings reveal pervasive differences between intuitive and analytic individuals at various levels of cognitive functioning.

The present results are also unique in the sense that they illustrate the everyday consequences of metacognitive differences as a function of analytic thinking. Namely, less analytic people are not only less effective at metacognitive monitoring when given a reasoning task, but they may also be less accurate at self-reporting their relative level of analytic thinking. It may be the case that this metacognitive advantage for analytic individuals may be part of the reason why analytic thinking is associated with a wide range of important psychological factors such as morality, religiosity, and creativity (see Pennycook, Fugelsang, & Koehler, 2015a). Thinking about how one thinks may influence how and what people think about. Indeed, our results suggest that part of the reason why debates in areas such as politics are so futile is because those for which overconfidence is the most consequential (i.e., those who need the most correcting, due to their low level of analyticity) are the least likely to recognize their overconfidence. Those most likely to be biased are also the least likely to recognize their bias.

Author Note

This research was funded by Natural Sciences and Engineering Research Council (NSERC), in the form of Discovery Grants (J.F. and D.K.) and an Alexander Graham Bell Canada Graduate Scholarship (G.P). Correspondence regarding this manuscript should be addressed to Gordon Pennycook, Department of Psychology, Yale University, 2 Hillhouse Ave, New Haven, CT, USA, 06520-8205 or by email: gordon.pennycook@yale.edu.

References

- Atir, S., Rosenzweig, E., & Dunning, D. (2015). When Knowledge Knows No Bounds: Self-Perceived Expertise Predicts Claims of Impossible Knowledge. *Psychological Science*, 0956797615588195-. <https://doi.org/10.1177/0956797615588195>
- Banks, A. P., & Hope, C. (2014). Heuristic and analytic processes in reasoning: An event-related potential study of belief bias. *Psychophysiology*, 51(3), 290–297. <https://doi.org/10.1111/psyp.12169>
- Bonner, C., & Newell, B. R. (2010). In conflict with ourselves? An investigation of heuristic and analytic processes in decision making. *Memory & Cognition*, 38(2), 186–196. <https://doi.org/10.3758/MC.38.2.186>
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42(1), 116–131. <https://doi.org/10.1037/0022-3514.42.1.116>
- Cacioppo, J. T., Petty, R. E., Feinstein, J. a., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, 119(2), 197–253. <https://doi.org/10.1037/0033-2909.119.2.197>
- Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition.: EBSCOhost. *Journal of Personality Assessment*. https://doi.org/10.1207/s15327752jpa4803_13
- Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: A process dissociation approach. *Journal of Personality and Social Psychology*, 104(2), 216–235. <https://doi.org/10.1037/a0031021>
- Critcher, C. R., & Dunning, D. (2009). How chronic self-views influence (and mislead) self-assessments of task performance: Self-views shape bottom-up experiences with the task. *Journal of Personality and Social Psychology*, 97(6), 931–945.

<https://doi.org/10.1037/a0017452>

De Neys, W. (2012). Bias and Conflict: A Case for Logical Intuitions. *Perspectives on Psychological Science*, 7(1), 28–38. <https://doi.org/10.1177/1745691611429354>

De Neys, W. (2014). Conflict detection, dual processes, and logical intuitions: Some clarifications. *Thinking & Reasoning*, 20(2), 169–187.

<https://doi.org/10.1080/13546783.2013.854725>

de Neys, W., Cromheeke, S., & Osman, M. (2011). Biased but in doubt: Conflict and decision confidence. *PLoS ONE*, 6(1). <https://doi.org/10.1371/journal.pone.0015954>

De Neys, W., & Franssens, S. (2009). Belief inhibition during thinking: Not always winning but at least taking part. *Cognition*, 113(1), 45–61.

<https://doi.org/10.1016/j.cognition.2009.07.009>

De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition*, 106(3), 1248–1299. <https://doi.org/10.1016/j.cognition.2007.06.002>

De Neys, W., Moyens, E., & Vansteenwegen, D. (2010). Feeling we're biased: autonomic arousal and reasoning conflict. *Cognitive, Affective & Behavioral Neuroscience*, 10(2), 208–216. <https://doi.org/10.3758/CABN.10.2.208>

De Neys, W., Rossi, S., & Houdé, O. (2013). Bats, balls, and substitution sensitivity: cognitive misers are no happy fools. *Psychonomic Bulletin & Review*, 20(2), 269–73.

<https://doi.org/10.3758/s13423-013-0384-5>

De Neys, W., Vartanian, O., & Goel, V. (2008). Smarter Than We Think. *Psychological Science*, 19(5), 483–489.

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, 12(3), 83–87.

<https://doi.org/10.1111/1467-8721.01235>

Ehrlinger, J., & Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of Personality and Social Psychology*, *84*(1), 5–17.

<https://doi.org/10.1037/0022-3514.84.1.5>

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *Journal of Personality and Social Psychology*, *71*(2), 390–405. <https://doi.org/10.1037/0022-3514.71.2.390>

Evans, J., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–241.

<https://doi.org/10.1177/1745691612460685>

Fleischhauer, M., Enge, S., Brocke, B., Ullrich, J., Strobel, A., & Strobel, A. (2010). Same or different? Clarifying the relationship of need for cognition to personality and intelligence. *Personality and Social Psychology Bulletin*, *36*(1), 82–96.

<https://doi.org/10.1177/0146167209351886>

Franssens, S., & De Neys, W. (2009). The effortless nature of conflict detection during thinking. *Thinking & Reasoning*, *15*(2), 105–128. <https://doi.org/10.1080/13546780802711185>

Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, *19*(4), 25–42. <https://doi.org/10.1257/089533005775196732>

Gauthier, K. J., Christopher, A. N., Walter, M. I., Mourad, R., & Marek, P. (2006). Religiosity, religious doubt, and the need for cognition: Their interactive relationship with life satisfaction. *Journal of Happiness Studies*, *7*(2), 139–154. <https://doi.org/10.1007/s10902-005-1916-0>

Johnson, E. D., Tubau, E., & De Neys, W. (2016). The Doubting System 1: Evidence for

automatic substitution sensitivity. *Acta Psychologica*, *164*, 56–64.

<https://doi.org/10.1016/j.actpsy.2015.12.008>

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*, 1121–1134. Retrieved from

<http://psycnet.apa.org/journals/psp/77/6/1121>

Mata, A., Ferreira, M. B., & Sherman, S. J. (2013). The metacognitive advantage of deliberative thinkers: A dual-process perspective on overconfidence. *Journal of Personality and Social Psychology*, *105*(3), 353–373. <https://doi.org/10.1037/a0033640>

Mata, A., Fiedler, K., Ferreira, M. B., & Almeida, T. (2013). Reasoning about others' reasoning. *Journal of Experimental Social Psychology*, *49*(3), 486–491.

<https://doi.org/10.1016/j.jesp.2013.01.010>

Noori, M. (2016). Cognitive reflection as a predictor of susceptibility to behavioral anomalies, *11*(1), 114–120.

Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology*, *76*(6), 972–987. <https://doi.org/10.1037/0022-3514.76.6.972>

Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2014). Cognitive style and religiosity: the role of conflict detection. *Memory & Cognition*, *42*(1), 1–10.

<https://doi.org/10.3758/s13421-013-0340-7>

Pennycook, G., Cheyne, J. A., Koehler, D. J., & Fugelsang, J. a. (2015). Is the cognitive reflection test a measure of both reflection and intuition? *Behavior Research Methods*,

(MARCH). <https://doi.org/10.3758/s13428-015-0576-1>

Pennycook, G., Cheyne, J. A., Koehler, D. J., & Fugelsang, J. A. (2016). Is the cognitive reflection test a measure of both reflection and intuition? *Behavior Research Methods*, *48*(1). <https://doi.org/10.3758/s13428-015-0576-1>

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015a). Everyday Consequences of Analytic Thinking. *Current Directions in Psychological Science*, *24*(6), 425–432. <https://doi.org/10.1177/0963721415604610>

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). Everyday Consequences of Analytic Thinking. *Current Directions in Psychological Science*, *24*(6). <https://doi.org/10.1177/0963721415604610>

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015b). What makes us think? A three-stage dual-process model of analytic engagement. *Cognitive Psychology*, *80*, 34–72. <https://doi.org/10.1016/j.cogpsych.2015.05.001>

Pennycook, G., & Ross, R. M. (2016). Commentary on: Cognitive reflection vs. calculation in decision making. *Frontiers in Psychology*, *7*, 9. <https://doi.org/10.3389/fpsyg.2015.00532>

Petty, R. E., Brinol, P., Loersch, C., & McCaslin, M. J. (2009). The need for cognition. In M. R. L. & R. H. Hoyle (Ed.), *Handbook of individual differences in social behavior* (pp. 318–329). New York: Guilford.

Sargent, M. J. (2004). Less thought, more punishment: need for cognition predicts support for punitive responses to crime. *Personality and Social Psychology Bulletin*, *30*(11), 1485–1493. <https://doi.org/10.1177/0146167204264481>

Slooman, S. A. (2014). Two Systems of Reasoning: An Update, 69–79.

Stanovich, K. E. (2012). On the Distinction Between Rationality and Intelligence: Implications

for Understanding Individual Differences in Reasoning. *The Oxford Handbook of Thinking and Reasoning*, 433–455.

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General*, *127*(2), 161–188. <https://doi.org/10.1037/0096-3445.127.2.161>

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: implications for the rationality debate? *The Behavioral and Brain Sciences*, *23*(5), 645-665-726. <https://doi.org/10.1017/S0140525X00003435>

Svedholm, A. M., & Lindeman, M. (2013). The separate roles of the reflective mind and involuntary inhibitory control in gatekeeping paranormal beliefs and the underlying intuitive confusions. *British Journal of Psychology*, *104*(3), 303–319. <https://doi.org/10.1111/j.2044-8295.2012.02118.x>

Thompson, V. A., & Johnson, S. C. (2014). Conflict, metacognition, and analytic thinking. *Thinking & Reasoning*, *20*(January 2015), 215–244. <https://doi.org/10.1080/13546783.2013.869763>

Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology*, *63*(3), 107–140. <https://doi.org/10.1016/j.cogpsych.2011.06.001>

Thompson, V. A., Turner, J. A. P., Pennycook, G., Ball, L. J., Brack, H., Ophir, Y., & Ackerman, R. (2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition*, *128*(2), 237–251. <https://doi.org/10.1016/j.cognition.2012.09.012>

Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive

reflection test. *Judgment and Decision Making*, 11(1), 99–113.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning*, 20(2), 147–168. <https://doi.org/10.1080/13546783.2013.844729>

Toplak, M., West, R., & Stanovich, K. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, 39(7), 1275–1289. <https://doi.org/10.3758/s13421-011-0104-1>

Travers, E., Rolison, J. J., & Feeney, A. (2016). The time course of conflict on the Cognitive Reflection Test. *Cognition*, 150, 109–118. <https://doi.org/http://dx.doi.org/10.1016/j.cognition.2016.01.015>